

# Fungibility and the Flypaper Effect of Project Aid: Micro-Evidence for Vietnam

Dominique van de Walle and Ren Mu<sup>\*</sup>

Research Development Group, World Bank

## Abstract

While most economists assume that aid is fungible, most aid donors behave as if it is not. Recipient government responses to development project aid are studied in the context of a specific World Bank-financed project. We estimate the impact of a rural road rehabilitation project in Vietnam on the kilometers of roads actually rehabilitated and built. Using local-level survey data collected for this purpose, we test whether the evidence supports the standard economic argument that there will be little or no impact on rural roads rehabilitated, given fungibility. We find evidence that, although project aid impacts on rehabilitated road kilometers were less than intended, more roads were built in project areas. Our results suggest that there was fungibility within the sector, but that aid largely stuck to that sector.

JEL *classification*: O19; O12; F35;

*Key words*: Impact evaluation, fungibility of aid, flypaper effect, rural roads, Vietnam

World Bank Policy Research Working Paper 4133, February 2007

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.*

---

<sup>\*</sup> We are grateful to Vu Tuan Anh, Christina Malmberg-Calvo, Dorothyjean Cratty, Binh Nguyen and Martin Ravallion for their encouragement, help and support. We received useful comments on this paper from Jean-Marie Balland, Kathleen Beegle, Michael Clemens, Hai-Anh Dang, Jishnu Das, Quy-Toan Do, John Giles, Elizabeth King, Martin Ravallion, participants at the 2004 NEUDC meetings in Montreal and at the 2006 “Effectiveness of Aid to Poor Countries” conference in Namur, and the journal’s two anonymous referees and editor. We gratefully acknowledge funding support from CIDA (TF034859), the World Bank’s East Asia and Pacific Transport Unit and the Poverty and Social Unit in the Vietnam Resident Mission Office, and the DFID trust fund for Poverty Analysis and Policy Advice. Address for correspondence: Dominique van de Walle, World Bank, 1818 H ST, NW, Washington, DC 20433. tel: 202-473-7935, [dvandewalle@worldbank.org](mailto:dvandewalle@worldbank.org).

## 1. Introduction

It is well recognized that the degree of displacement of local resources by external aid has important implications for assessing aid effectiveness. When aid that is ostensibly tied to a specific intervention is in fact fungible, it becomes extremely difficult to determine what exactly the aid has financed and hence what development impact it has had on the ground. Yet prevailing views on the extent of fungibility vary greatly. At the risk of over-generalizing, most economists appear to think that fungibility is the norm, while most aid donors behave as though there is no fungibility.

Development assistance continues to be primarily project-based, particularly for infrastructure. The possibility of fungibility is routinely ignored in project work. Project staff often put enormous effort into project selection and appraisal. If fungibility is indeed the reality, then these efforts are wasted.<sup>1</sup> This is not to deny the importance of ex-ante appraisal and design issues, but to question whether they should be linked to project aid.

This paper studies the degree to which road project funding actually ends up funding what the donor intended. We focus on the first World Bank-financed Rural Transport Project (RTP1) in Vietnam — a project launched in 1997 and aiming to rehabilitate 5,000 kilometers of district and commune level roads in 18 provinces.<sup>2</sup>

Our starting hypothesis is that there is full fungibility. The very idea of aid earmarked to rehabilitate rural roads runs against the simple economic logic that resources are fungible. Recipient governments have their own preferences over spending. As long as the amount of aid is less than they would have spent, they will cut their own funding of roads in the project areas in response to the external aid. One still expects to see an income effect of the extra aid on road (and other) spending, but one should not see a differential impact on project areas versus the non-project areas used for comparison purposes. Against this hypothesis, there is empirical evidence of “flypaper effects” whereby central grants to local governments stimulate much higher local spending of the intended type than would an increase in community income. Thus the money “sticks,” as flies on flypaper.

---

<sup>1</sup> There may be capacity-building effects that foster better public decision-making elsewhere.

<sup>2</sup> The Vietnam Rural Transport Project I, see World Bank (1996) for details.

In the case of this Vietnam rural roads project we ask whether there is evidence to support the standard economic argument on fungibility, or whether the evidence supports the alternative assumption made by aid donors, that donors get what they paid for.

An initial casual look at our project evaluation data suggests partial fungibility. We find about half the expected difference between project and non-project comparison communes in the length of roads undergoing rehabilitation during the project period. Taking a simple mean difference between participating and non-participating communes shows a 2.5 kilometers cumulative average difference in favor of the project. Yet, independent administrative data reports that an average of 4.6 kilometers per commune was rehabilitated under the project. Project supervision and other evidence confirm that the project was implemented as planned.

This appears to indicate partial fungibility, representing about half of the resources received. But on closer inspection, the data further suggest that the project communes built significantly more kilometers of new roads than did the non-project communes. This suggests the possibility of a sectoral flypaper effect of the aid within the road sector. Yes, displacement occurred in that less rehabilitation occurred than the aid had (ostensibly) financed. However, the aid stuck to the sector with the donor getting more built roads.

As often argued in the literature, empirical evidence for the flypaper effect may be due to model misspecification and entirely spurious (Hines and Thaler, 1995, review the arguments). An alternative explanation for our observed difference in rural road length is that we are not using an appropriate counterfactual. In judging the impact of the project we may need to deal with the possibility of selection bias. What looks like a flypaper effect of project aid may in fact be the outcome of biases in the empirical methods.

While rural roads have been extensively championed as poverty alleviation instruments by aid donors, rigorous impact assessments have been rare. Such assessments require appropriate controls, and must allow for the factors that influence both program placement and outcomes. We use a panel data set (collected for the purpose of conducting such an impact evaluation) of communes and households within project and non-project areas. The dataset comprises a baseline and post-project follow-up data. The data allow us to more rigorously test our hypothesis of full fungibility and no flypaper effect of the project aid on balancing the observable covariates of project and comparison areas.

Our impact estimates controlling for selective placement of the aid project suggest the existence of a sectoral flypaper effect in that part of the project funds were diverted from rehabilitation to the building of new roads but stayed within the roads sector. Project communes ignored the donor stipulation that they not build new roads. Instead, they imposed their own priorities and diverted some of the funding to priority road building. We also find that, in general, the quality of rehabilitated roads improved in the project communes. This entailed a switch from the donors preferred technology, namely earth road rehabilitation.

The paper first briefly reviews the literature on aid fungibility and the flypaper effect. This is followed by a description of the project being evaluated and the setting. Section 4 describes the survey instrument and some very preliminary results, while section 5 lays out the methods used. Section 6 discusses the main results and other evidence. Section 7 concludes.

## **2. Development aid fungibility and the flypaper effect**

A number of studies have investigated the degree to which donors' earmarked development or project aid substitutes for, rather than supplements, local spending intended for that purpose. Fungibility — whereby categorical aid substitutes for government funding by being diverted to spending on other sectors or being used to reduce taxes — tends to be difficult to detect and to identify empirically.

Empirical investigations of aid fungibility have attempted to identify government responses from the variance in spending across countries or from within country changes in spending over time. For example, many past studies have used pooled cross-country time series data to examine the impact of tied aid on government spending across sectors (for example, Feyzioglu et al., 1998; World Bank, 1998). Another set of studies have used country specific time series of macro aggregates to explore the effect of categorical aid on the public sector's expenditure mix (Pack and Pack, 1990 and 1993). Government responses to targeted aid are found to vary quite a lot. But with notable exceptions, a significant amount of fungibility tends to be found in this strand of the literature.

A related public finance literature investigates the flypaper effect. This refers to the observed phenomenon that the expenditure stimulated by central to local government

grants is much larger than an equivalent increase in income. Economists tend to regard this finding as anomalous in that it contradicts the standard theoretical result from the unitary model of local government decision making. In this model the local authority is thought of as a single representative consumer, or median voter, who will treat the aid as ordinary income. In this case, the flypaper effect suggests irrationality. However, a number of theoretical papers have constructed models under which a flypaper effect can exist in reality.

The recipients of aid — provinces and communes in Vietnam — are heterogeneous entities. Then, it is no longer clear that there is anything “anomalous” about the flypaper effect. Indeed, Roemer and Silvestre (2002) show that when outcomes are determined by an electoral process, the general presumption will be some degree of flypaper effect. They show that the complete absence of a flypaper effect requires rather special conditions: namely that incomes are essentially shared the same way as the tax burden. The relevance of this kind of theoretical model to Vietnam, where local governmental processes are not electoral, is questionable. However, the general point of Roemer and Silvestre (2002) remains, namely that collective decision making processes with heterogeneous preferences within recipient governments can sometimes generate flypaper effects.

A large developed country literature has grown around investigating and explaining this common empirical finding (for an overview see Hines and Thaler, 1995). Similarly to the fungibility literature, the empirical work largely relies on cross-sectional or pooled cross-sectional time series data on grants to (in this case within country) localities such as states or school districts.

There have been few applications of this literature to developing countries. An exception, though in a quite different context, is Jacoby’s (2002) study of whether child consumption increases commensurately to food intake as a result of participation in a school feeding program in the Philippines. In common with Jacoby (2002), this paper uses a quasi-experimental design. This differs in a number of ways from the past work on fungibility and the flypaper effect in the context of grants to local governments.

The first major difference concerns the level of aggregation. There appears to be little evidence based on actual development projects, which are often implemented at the

local level, under increasingly more decentralized fiscal arrangements. At the level of aggregation at which such studies are typically conducted, a lot of the variance in project aid, and responses to it, will be lost. For one, the situation in the absence of aid is not observed. This undoubtedly leads to a loss in the ability to identify the key parameters of interest. In contrast, since external aid projects can't be implemented everywhere, there will be places that don't get them. Observations on places without the project aid provide us with useful extra information that other studies have not had.

A second difference concerns the degree of heterogeneity that can be allowed for in our approach. Differences in government preferences will be difficult to take into account in cross-country regressions. One has a better hope of identifying the displacement of aid using country-specific data. But country studies have also tended to focus only on central government behavior. A finding of no fungibility at the central level could well be compatible with widespread displacement at lower government levels. Our approach allows us to explore whether displacement is found in local government behavior — at the level where the key decisions are made. Against that, our approach has the drawback that we may be unable to detect displacement that occurs at the highest level of government. In the case of the rural roads project studied here, displacement could occur either at the local or higher levels. Vietnam's central government may well reduce its budget allocation to the Ministry of Transport (MOT) as a response to the World Bank's project funding. We test for displacement at lower levels of government, but are unable to do so at the central level.

Funding from intermediate levels of government (district or province) that had been slated for road improvements in the project communes may get diverted to communes not benefiting from the donor-financed project, or to other basic commune infrastructure that they also typically fund. We think this is the most likely way that displacement occurs in this setting. In principle, the local commune authorities could divert resources that would have been spent on roads to other commune-level needs or to reduce the tax burden, for example by reducing residents' labor burden.<sup>3</sup> However,

---

<sup>3</sup> Able-bodied adults in Vietnam are expected to contribute 10 labor days yearly (or the cash equivalent) to finance small commune infrastructure construction and maintenance.

communes are typically resource poor, relying on the provincial or district governments to undertake road rehabilitation or other basic infrastructure construction (Rao et al., 1999). Furthermore, under the project, provinces retained substantial control of project funding. Thus, displacement is unlikely to have been significant at the commune level.

Behavioral responses by aid recipients could also result in road spending being used differently in project areas. For example, when labor inputs are supplied locally and are fixed, project funds may encourage a switch in local road rehabilitation efforts away from more labor intensive techniques (more sealed roads, for example). The outcome may be fewer rehabilitated kilometers but higher quality rehabilitation. There can be a tension with donor preferences here given that the development aid community has emphasized labor intensive earth road rehabilitation methods.

It should be noted that even with full fungibility, one may still see a positive impact on road spending. Fungibility means that one is free to spend on whatever one likes. A pure income effect can still be expected on the demand for all normal goods including roads. This poses a problem for measuring fungibility (Khilji and Zampelli, 1994). Because of a potential income effect of the fungible component of aid funding, we cannot identify the precise amount of fungibility. The joint outcome of fungibility and an income effect can be called “displacement.”<sup>4</sup> If resources are not fungible and a flypaper effect is found instead, the issue of an income effect does not arise.

### **3. The project intervention and participant selection**

The Vietnam Rural Transport Project 1 (RTP1) is a rural roads rehabilitation project aiming to link commune centers to markets and so reduce poverty (World Bank 1996). Communes are the lowest administrative unit in Vietnam, above which are districts and, above them, provinces. Communes have an average of about 1,300 households living in 10 or so hamlets (or villages).<sup>5</sup>

The project was launched in 1997 for implementation in 18 poor provinces over 4

---

<sup>4</sup> In order to quantitatively isolate fungibility from the income effect we would need an evaluation design whereby some communes got an allocation of untied funds equal to that received under RTP1. One could then evaluate the effect of that allocation on road spending in those communes and add that amount to the displaced spending to get at the degree of actual fungibility.

<sup>5</sup> This is based on the 200 communes in our survey, described in the next section.

years, at a cost of about \$61 million. It aimed to rehabilitate 3,500 kilometers of district roads and 1,500 kilometers of communal roads. In each participating province, road links were identified for rehabilitation through least cost techniques.<sup>6</sup> The project stipulated that no new roads should be built. However, eligible existing roads were often impassable by motor vehicle for much of the year.

As noted, RTP1 funds were earmarked for the rehabilitation of commune and district level (or rural) roads. These roads are normally the responsibility of commune and district governments who typically lack the resources for anything other than routine maintenance, which is implemented by commune residents. When funding from the central government or international donors does become available, the provincial authorities channel the funds to the targeted communes. The project is implemented through the central government's MOT, which chose provinces for participation. Its provincial offices were then in charge of picking communes and road links to be rehabilitated, as well as appropriate contractors following a competitive bidding process.

Road link selection is in theory based on a number of conditions imposed by the project: average investment costs being no more than \$15,000 per km and the population served being at least 300 people per kilometer. In an effort to extend project benefits to low density, mountainous areas with concentrations of ethnic minority populations, 20 percent of each province's rehabilitation funds could be set aside for roads not justified under the population and cost criteria.

#### **4. Data and preliminary results**

We use a data set that was created specifically for evaluating the impact on living standards of the rural roads rehabilitated under RTP1. The "Survey of Impacts of Rural Roads in Vietnam" (SIRRV) is a panel data set of pre-project baseline and post-project data for both project ("treatment") and non-project ("comparison") areas.<sup>7</sup> The data were collected in six of the 18 provinces included in RTP1. Two provinces were randomly picked from the participating provinces in the country's north (Lao Cai and Thai Nguyen), center (Nghe An and Kon Tum) and south (Binh Thuan and Tra Vinh). The first author

---

<sup>6</sup> Least cost techniques refer to the minimum-cost engineering solution that ensures a minimum level of motorized passability.

<sup>7</sup> For a more detailed description of the SIRRV see van de Walle (2007).



designed and supervised the data collection and the implementation was done by the Institute of Economics in Hanoi.

In each of the six provinces, samples of project and non-project communes were drawn for a total of 200 surveyed communes. Road links covered by the project all pass through communes, and a majority link up commune centers—where facilities and services are located—with the road network. Data are often, and more easily, collected at the commune level in Vietnam. For these reasons, a road's zone of influence is defined as the commune through which the road passes. The project communes were randomly selected from lists of all communes with proposed projects in each province. A list was then drawn up of all remaining communes in districts with proposed sub-projects from which a random sample of non-project communes was drawn. Comparison areas chosen this way should share many of the same characteristics as the project areas. However, we cannot be confident that they are a good comparison group on a priori grounds. For this reason, we will use matching techniques to test the selection of comparison groups.

A detailed commune-level data base was created in part by drawing on annually collected records—both current and retrospective—at the commune level and augmenting this information with other data. A household questionnaire was also administered to 15 households in each sampled commune.<sup>8</sup>

The baseline data collection began in June 1997. Subsequent rounds followed at two year intervals, in the summers of 1999 and 2001, tracking the implementation process followed initially as closely as possible. Surveying of districts is timed across months to coincide with the schedule followed during the baseline. In this paper, we use only the baseline and 2001 waves.

We also use a project level database constructed from provincial and central records. This provides information on project location, timing, costs, the condition of the road pre- and post-project, whether bridges were rehabilitated, and kilometers rehabilitated. Some of these data were not carefully recorded, or are missing entirely. However, the aggregates do accord well with kilometers recorded by the center; the means

---

<sup>8</sup> Given constraints, no attempt was made to measure a household level welfare indicator such as consumption. However, the household survey was designed in view of combining it with the nationally representative 1998 Vietnam Living Standards Survey (VLSS) to predict baseline consumption expenditures for SIRR households (van de Walle, 2006).

correspond quite closely, but the commune-level numbers are more suspect. Errors in assigning the project funds appear to average out.

Much of the analysis in this paper focuses on a section of the commune questionnaire that asks about the kilometers of roads that were rehabilitated and built during the last two years. A potential data issue is that routine maintenance appears to have been recorded as rehabilitation for some communes.<sup>9</sup> We suspect that this is the case when the kilometers rehabilitated are unusually high and equal, or very nearly equal, the commune's entire road or earth road network, the costs are low and the work has been entirely implemented by villagers' labor contributions, with added funding from the commune only. The incidence of such cases is more common in the baseline which could indicate that road maintenance is declining over the period as more funds get allocated to rehabilitation. Or it could reflect a change in how the question was interpreted by the interviewers over time.

This potential data problem may affect selection into the project and project impacts. It can be conjectured that the demand for rehabilitation will tend to be low in places where maintenance is well done and that the project will be allocated to areas with a poor record in maintenance. However, in testing for this we found that such communes were no more (or less) likely to participate in the project. We set those pre-project rehabilitated kilometers to zero on the assumption that although roads were maintained, none were rehabilitated. This data issue may be more problematic for calculating impacts. So, we exclude communes with very large outliers during the project period. This leaves us with 98 project and 96 non-project communes.

In theory, for full fungibility to be possible, the amount allocated needs to be less than or equal to the amount being spent anyway. The recipient then spends what had been intended and the aid is treated as general budget support. We find that for the bulk of project communes — 82 out of 98 — the kilometers actually rehabilitated during the project period were greater than or equal to the kilometers allocated by RTP1. Superficially then, these simple descriptive statistics establish that full fungibility was at

---

<sup>9</sup> Routine maintenance refers to work that is not intended to cover major repairs or upgrades, and is required annually or periodically on all roads, rendering rehabilitation unnecessary. Rehabilitation is done on poor and very poor roads that have not received proper maintenance and might include regravelling, slope

least possible in most communes.<sup>10</sup>

The simplest method for testing fungibility with these data is to regress the actual kilometers rehabilitated during the project period on the amounts allocated by RTP1. This gives a regression coefficient of 0.55 ( $t=2.13$ ), which is well below, but not statistically different from, unity. The constant term is 12.30 ( $t=9.35$ ), reflecting the road rehabilitation that would be done in the absence of the project (and again more than the allocated project amount). When we test built roads in the same way, there is little correlation (the coefficient is 0.24 with  $t=1.2$ ). Finally, looking at total built and rehabilitated kilometers, the coefficient is 0.79 ( $t=1.94$ ), which is closer to one although not significantly different from zero at the 5 percent level. Indeed, we can't reject either the null hypothesis that it is zero or that it is one.

The above regressions are suggestive of partial fungibility of the project aid and a road-sector flypaper effect. However, as mentioned previously, we have concerns about the quality of the data on kilometers allocated to communes by RTP1 from the project database. The means appear reliable and reflect donor and government decisions but the commune allocations are more questionable. Possibly we are underestimating impacts due to attenuation biases. In addition, there is an endogeneity concern. This can be thought of as having two aspects: the assignment of which communes get the project, and the amounts allocated given the assignment across communes.

In the light of these observations, we will treat the intervention as a zero/one dummy variable, rather than by the kilometers allocated as recorded in the project database, and we will model this dummy as a function of the observed covariates of program placement. So, we aim to estimate the most common evaluation parameter, the average treatment effect on the treated (ATT) as given by the roads built and rehabilitated by the project less the outcomes under the counterfactual of not receiving the project aid. We then compare this to the independent administrative data on the aggregate allocation.

Despite the non-project communes being originally selected to be “similar” to the project communes, one can expect differences between them. For example, this could

---

protection and drainage repairs. Building refers to completely new road construction with new vehicle access. The SIRR question on road building clearly notes to exclude “rehabilitation.”

<sup>10</sup> Later, after discussing our estimation methods, we ascertain that this holds for almost all communes when we predict the counterfactual kilometers of rehabilitated roads in the absence of the project.

come from the need for rehabilitation. The extent of rehabilitation that is implemented is likely to be highly related to initial road infrastructure conditions. We want to control for this and other factors influencing project placement and potentially correlated with the amount of new activity in the local road sector.

## **5. Methods for addressing selection bias**

Impact estimates can be sensitive to the estimator chosen (see, for example, Smith and Todd, 2005). For this reason, we calculate a number of different estimators. In all cases, the “outcome variable” is the number of kilometers of roads rehabilitated or built during the project period. We start with a conventional double-difference (DD) estimator. This can be obtained from an OLS regression where the dependent variable is the change in measured road works between the pre-intervention baseline and the post-project date and the only explanatory variable is a dummy for whether the commune participated in the project. This assumes that the project is placed as a function of initial conditions that take the form of an additive time-invariant error term. Once this has been differenced out, any differences in outcomes between the treated and non-treated communes can be attributed to the project.

However, the ordinary DD estimator does not allow for initial conditions that also influence the subsequent changes over time. One way to allow for these is to interact them with time in the date-specific outcome levels by including them in the DD regression. However, this method is known to have a number of drawbacks. One is that it imposes arbitrary functional form restrictions on the outcome equations. Another is that it uses data on all non-treated communes even though some may be very different from treated communes, raising questions about the robustness of relying on functional form to extrapolate outside the common support. These drawbacks can be avoided by combining double differencing with propensity score matching (PSM) to select ideal comparison communes from among the sampled non-project communes.<sup>11</sup>

We use two alternative methods based on propensity scores (PS) to construct an appropriate counterfactual. In the first method, a DD estimator is combined with matching on PS; in the second method, a DD estimator is weighted by the PS.

---

<sup>11</sup> For a discussion of the differences between PSM and other methods, see Ravallion (2006).

To see this more clearly, let  $P(Z) = \Pr(D = 1 | Z)$  denote a commune's probability of having a road link project (its propensity score), where  $Z$  is a vector of observed commune characteristics that determine selection into the treatment group. Rosenbaum and Rubin (1983) show that units (here communes) with similar PS also have a similar distribution of covariates  $Z$ . Under the assumption that, after conditioning on the PS, outcome means are independent of treatment, PSM allows us to form a comparison group for each project commune by choosing the non-project communes with the closest PS, or assigning different weights to the non-project communes based on their PS. The estimated ATT is  $(\sum_{i \in T} Y_i - \sum_{j \in C} W_{ij} Y_j) / N_T$ , where  $T$  denotes the set of ("treated") project communes,  $C$  denotes the set of comparison communes,  $N_T$  is the number of project communes and  $W_{ij}$  is the associated weight given to the  $j$ th commune in making a comparison with the  $i$ th project commune.

How many non-project communes to include in the comparison group and how to assign weights to them are important practical questions in implementing PSM. One popular option is to use nearest-neighbor matching. However, recent theoretical work shows that the extreme non-smoothness of nearest-neighbor matching means that the conventional bootstrapping method is inappropriate for estimating the standard errors (Abadie and Imbens, 2006a and 2006b). In order to get valid bootstrapped standard errors, we choose instead to apply nonparametric kernel matching in which all the non-participants are used as comparison communes and weights are assigned according to a kernel function of the predicted PS.<sup>12 13</sup>

The alternative method exploits the PS to achieve a balancing in the covariates following Hirano et al. (2003), who show that weighting the observations with the PS (and indirectly, the covariates) can create balance between treated and non-treated units in the

---

<sup>12</sup> The weights can be expressed as  $W_{ij} = \frac{G((P_j(Z) - P_i(Z)) / a_n)}{\sum_k G((P_k(Z) - P_i(Z)) / a_n)}$ , where  $G(\cdot)$  is a kernel function

and  $a_n$  a bandwidth parameter. In estimation we use the normal kernel function.

<sup>13</sup> We match on the log-odds ratio because project communes are over-sampled relative to their frequency in the population eligible for the project. Heckman and Todd (1995) show that matching on the odds ratio is robust to choice-based sampling.

weighted sample. The weighting estimator is obtained by a weighted least squares regression of  $Y_i = \alpha + \beta D_i + \varepsilon$ , where the weights are unity for the treated units and  $\hat{P}(Z)/1 - \hat{P}(Z)$  for the comparison units (Hirano and Imbens 2002; Imbens 2004).<sup>14</sup> Hirano et al. (2003) shows that this estimator is efficient.

These methods attempt to control for observed pre-project heterogeneities that may lead to selection into the project as well as influence outcomes and the growth path of the communes. We assume that any selection bias due to unobserved factors is time-invariant and so can be differenced out. Accordingly, the DD PS-matched estimator is given by:  $(\sum_{i \in T} (Y_{i1} - Y_{i0}) - \sum_{j \in C} W_{ij} (Y_{j1} - Y_{j0})) / N_T$ . Similarly, the DD PS-weighted

estimator is given by regressing the change in outcome on the treatment indicator:

$\Delta Y_i = \alpha' + \beta' D_i + \xi_{it}$ , with weights equal to unity for the treated and  $\hat{P}(Z)/1 - \hat{P}(Z)$  for the comparison communes.

As a further test for robustness, we use the PS to create two additional samples over which we calculate the estimates. In one we trim three non-project communes with PS lower than the lowest treatment's PS. We refer to this as 'treatment support.' In the second we further trim the six project communes with PS that are higher than any of the non-project communes' thus limiting the sample to the region of common support. This provides tighter matching but also results in a loss in our ability to make inferences about the population of project villages. We report these estimates as a test of the sensitivity of the full sample results.

Our estimates will be biased if there are omitted variables that influence both placement and the changes in outcomes. Selection bias on unobservables can never be fully ruled out in any observational study. While an array of initial conditions, including the initial state of infrastructure, will be controlled for in the logit model used to calculate the PS, we can never rule out the possibility of omitted initial conditions, or changes over time in those variables that are correlated with placement or outcomes. We will comment on possible omitted variables.

---

<sup>14</sup> If we wish to estimate the average treatment effect for the population, the weight is  $1/\hat{P}(Z)$  for the treated and  $1/(1 - \hat{P}(Z))$  for the non-treated (Hirano and Imbens, 2002).

## 6. Results

Table 1 shows means of kilometers built and rehabilitated during the three years prior to the project, and the four years of project duration, in both project and non-project communes as reported by the communes. The table also presents the breakdown of the types of road works, to which we will return. For the most part, the means for the project and non-project communes follow each other rather closely. Project areas have somewhat higher total road kilometers in both periods but the increment over time is about the same. Strikingly, the only significant difference is not for kilometers rehabilitated, but for kilometers built in the project communes between the start and end of the project.

### 6.1 Matching methods

Table 2 presents the logit model of commune participation in the road rehabilitation intervention.<sup>15</sup> The dependent variable takes the value one for communes with a project and zero for the others. The aim here is to control for all pre-determined variables that could affect participation in the project and are exogenous to it. We control for as many baseline commune level characteristics as are in our data, are not afflicted with missing observations, and are not too highly correlated among themselves.<sup>16</sup>

The pre-project condition of roads can be expected to be important in explaining why a commune was chosen for inclusion and subsequent impacts. Our measures of rural road density and the share of impassable earth roads should get at this indirectly. We are careful not to include any policy variables that are determined by higher levels of government also responsible for picking RTP1 participants — for example, whether a commune received funding outside for rehabilitation in recent years — since matching on variables that are jointly determined with project participation will lead to mismatching.

The logit contains no direct measures of political power or bureaucratic efficiency.

---

<sup>15</sup> The standard errors (here and for the regression based impact estimates presented shortly) allow for clustering at the district level. There are 29 districts.

<sup>16</sup> One potential concern is that there are too many variables in the logit, given the small number of observations. We tested robustness by dropping the 8 variables (population density, waterway, share of population in SOEs, share of perennial land, market, agricultural land affected by natural disaster, lower secondary school, and number of production organizations) with t-stats less than one. The pseudo  $R^2$  drops slightly to 0.23 but this made little difference to the variables remaining in the logit and had remarkably little effect on the impact estimates calculated using the resulting propensity scores.

We would argue that any differences in local government efficiency or influence are likely to be reflected in initial commune characteristics, which the logit does control for. And, they are unlikely to affect the project's impact on road kilometers rehabilitated. Once allocated to communes by the provincial management units (PMU), the road link projects are contracted out to bidding companies who are also chosen and supervised by the PMUs. Thus the commune authorities neither control nor oversee the road works. Nonetheless, it is possible that there is an unobservable factor that influences selection as well as the trajectory of the project impact on kilometers rehabilitated over time.

A number of explanatory variables are significant in Table 2. Participation is more likely in communes with higher total population (controlling for density) and a larger ethnic minority population — both official project selection criteria as discussed in Section 3. On the other hand, there is no sign that mountainous areas had any impact on participation controlling for their share of minority population. Higher school enrollments, a higher share of the population working for private enterprises, access to Agricultural Bank credit, the availability of passenger transport, a national road in the commune, distance to the province center and a higher road density had significant negative effects. A railway passing through the commune improved the likelihood of being selected to participate. Finally, communes in all provinces were more likely to get a project than those in Tra Vinh.

Figure 1 shows that there is imperfect overlap in the densities of the estimated propensity scores for project and non-project communes. This suggests that it will be important to use matching or re-weighting for selecting comparison communes and in determining the evaluation sample with support.

## **6.2 *Double difference estimates***

Table 3 presents the various DD estimates of the increment to rehabilitated road kilometers due to the project. These represent the mean change in rehabilitated kilometers in the project communes over and above the change in the non-project communes. The table also gives DD estimates of the increments to newly built and total road works. As detailed in Section 5, we give three estimators, namely a simple DD (first column), a PS kernel-matched DD (second), and a PS-weighted DD (third). Each of these is in turn



presented for three samples — namely the total sample, the sample minus three non-treatment communes with PS lower than any of those of the treatment communes (treatment support) and the sample further trimmed of six treatment communes to ensure common support.

Table 3 also gives the mean per commune kilometers allocated by RTP1, as recorded in the project data base. For our sample, RTP1 rehabilitated an average of just over four and a half kilometers per commune. Finally the last rows of Table 3 provide the p-values for two statistical tests of no fungibility. The first tests whether the rehabilitated kilometers are equal to that allocated by RTP1. The second tests whether there was fungibility within the roads sector — namely, whether the total kilometers rehabilitated and built over the project period equal the kilometers allocated by RTP1.

We first use the kernel matching method to calculate the kilometers that project communes would have rehabilitated without RTP1. The mean counterfactual of roads rehabilitated is 11.79 compared to the project allocation of 4.58 kilometers. The predicted amount in the absence of the project was greater than the project allocation for all but three communes verifying that full fungibility was feasible.

The simple DD comparison of the mean incremental kilometers rehabilitated by all communes indicates a statistically insignificant difference of 2.5 kilometers by the end of the project period (Table 3). In general the estimates are qualitatively similar within each method when we trim the comparison and/or treatment samples outside the region of common support (whether or not we trim the treatment group). They range from 1.68 to 3.88 kilometers rehabilitated, based on the simple DD and the PS-weighted DD, respectively. But, with the exception of the PS-weighted DD estimates (significant at the 10% level for the total and treatment support samples) none of the DD estimates of impacts on rehabilitation are significantly different from zero. The PS-weighted estimates indicate an average of 15 percent displacement relative to the per commune average of 4.58 kilometers rehabilitated by RTP1.

As an aside, we also tried a more conventional estimation method, namely using a parametric regression for the change in roads rehabilitated or built that includes initial conditions interacted with time in the date-specific outcome levels. As noted in Section 5, this method makes a number of rather different assumptions, including on functional form

and the use of the full sample, and it cannot be assumed that it would give similar results to those in Table 3. In fact we find quite striking differences in the impact estimates by this method. Controlling for the same initial conditions as are in the logit in this way results in DD estimates that are consistently considerably lower than those in Table 3, and statistically insignificant. For example, the impact estimates of rehabilitated kilometers due to the project are -0.31 (s.e.=1.75), -0.27 (s.e.=1.77) and -0.11 (s.e.=1.78) for the total, treatment and common support samples, respectively. For built roads, the estimates range from 1.66 to 1.79 with only that on the common support sample significant at least at the 10% level; and none of the estimates for total kilometers are statistically significant. Given that matching or re-weighting the data using PS can claim to be the most general way of controlling for observable heterogeneity, we take the results in Table 3 to be preferable. However, given the popularity of the linear regression DD estimator, its seemingly poor performance in this case is notable.

Focusing on the results for rehabilitation in Table 3, there are signs of partial fungibility. Although we cannot reject the null hypothesis that there is no fungibility with respect to rehabilitated road kilometers, the amounts are appreciably less than were allocated suggesting some fungibility. By contrast, all the estimators indicate that project communes actually built significantly more new roads than the comparison communes. The PS-based estimates indicate a statistically significant net gain of two and a half to three kilometers of newly built roads due to the project. This suggests that the communes more than made up for any missing rehabilitation by building roads instead.

This is reflected too in statistically significant estimates of the total — rehabilitated and built — kilometers of road works attributable to the project. For our two PS-based estimates, the total amount is persistently larger than the project allocation, suggesting that the project not only supplemented planned commune spending but stimulated additional road spending as well. However, statistically speaking, the estimates are not significantly different from the total allocated by the project as indicated by the p-values.

Our results indicate that the project has had impact on total road kilometers. Overall, the aid resources appear to be quite sticky – the project money stayed within the road sector. However, within roads, there was some fungibility whereby some of the aid may have gone to the construction of new roads as opposed to the rehabilitation of

existing roads. Below we discuss other evidence that is supportive of this conclusion.

### **6.3 Additional support**

We expect to see a lower degree of fungibility in communes with initially low levels of spending on rehabilitation relative to their allocation. Thus, our impact estimates for road rehabilitation should be higher in such communes. Tables 4a and 4b present the same impacts as given in Table 3 re-estimated for the sample divided into the 53 project communes that rehabilitated less pre-project than the amount subsequently allocated by the project, and the 45 that rehabilitated more. As can be seen, the mean RTP1 allocation was larger for the group that rehabilitated less pre-project.<sup>17</sup>

The tables suggest a lower degree of fungibility for prior low spenders than for high pre-project rehabilitation spenders. The impact estimates for rehabilitated km are significant in Table 4a, but not significantly different from zero in Table 4b. The latter result is the same as for the sample as a whole. And as in Table 3, we cannot statistically reject the null hypothesis that the rehabilitated kilometers equal the allocated kilometers for either sample. Furthermore, the impact estimates of built kilometers attributable to the project are significantly different from zero for both groups, as are total road works.

Hence, the stipulation that the resources be used for rehabilitation was more binding in places that were implementing less rehabilitation prior to the project. Yet, although the degree of fungibility was affected commensurately, the project resources stuck to the sector regardless. The test that the total road works equal the amount allocated can not be rejected for the preferred PS-based estimates. For both commune samples, the evidence points to a flypaper effect.

Additional corroborating evidence comes from available SIRRV data on other construction carried out during the project period. If the resources allocated for road rehabilitation had been displaced to the non-road sector, we would expect this to be reflected in commune spending on other basic infrastructure that is typically funded by the provincial authorities who disbursed the road funding. The data on school, health center, commune center and office construction do not suggest that resources were diverted to these other uses. The DD estimates in Table 5 generally do not indicate differential

---

<sup>17</sup> Note also that the first group rehabilitated 16 km of roads on average during the project period while the second group rehabilitated 13 — both considerably more than the average RTP1 allocation.

impact between the project and comparison communes, with three exceptions. The PS-weighted DD suggest that 18 percent more project communes invested in their health centers, while the PS-matched DD suggest the 17 percent fewer carried out construction on commune cultural centers. (Both are significant only at the 10 percent level for the total and treatment support samples.) When we consider these investments jointly, we find that the PS-matched estimate on the total sample is significant at the 10 percent level — suggesting that 15% more project communes implemented construction of any kind. On balance however, this evidence also supports our conclusion that there is a flypaper effect on the road sector.

#### **6.4     *Impacts on road quality***

We have focused on total rehabilitated and built kilometers without distinguishing between the types of roads that were worked on. As noted earlier, Table 1 presents a before and after project comparison of the breakdown in the types of roads built and rehabilitated in project and non-project communes. The underlying time periods differ but the percentage shares over time are comparable. Compared to what the project areas were doing pre-project, there are clear signs of a switch away from the rehabilitation of motor vehicle impassable earth roads and towards the rehabilitation of both paved all-weather, and paved sometimes-impassable roads. Similarly, in building new roads, a compositional shift away from earth towards paved roads is apparent, relative to the pre-project period. Similar, though possibly less pronounced, patterns are found for non-project communes.

To examine this more rigorously, Table 6 presents DD estimates of the increments to different road types attributable to the project. Here too, we see signs of a shift, whereby project communes focused their rehabilitation efforts on paved all-weather, and paved sometimes-impassable roads, more so than non-project communes. All estimates, on all samples, support the shift to paved all-weather roads. The PS-matched and weighted estimates provide support for the shift to paved sometimes-impassable roads but also to greater rehabilitation of passable earth roads. However, in terms of road building, the project impact is felt primarily on the kilometers of earth all-weather roads. Hence, the shift away from motor-impassable earth towards paved all-weather roads seen for project communes in Table 1 also occurred in the comparison communes.

These results strongly suggest that the project caused a shift in road rehabilitation

efforts from impassable earth roads to better earth roads and to paved roads within project communes. These shifts were contrary to the revealed preferences of the aid donor for a lower rehabilitation standard as expressed in the official project documents.

## **7. Conclusions**

This paper has studied the issue of project aid displacement in the context of a specific aid-financed development project. We have estimated the impact of a World Bank-financed rural road rehabilitation project in Vietnam on the kilometers of roads actually rehabilitated and built by local government agencies.

From a theoretical perspective, the expected outcomes are ambiguous. Treating aid recipients as homogeneous entities, economic theory predicts that if project aid is less than or equal to what recipients intended to spend, then they will spend as planned and treat the aid as generalized budget support. The aid will be fully fungible. Against that, empirical evidence for flypaper effects is plentiful and theoretical contributions have shown how such effects can be generated when one allows for heterogeneity within aid-receiving units and collective decision making. Others have argued that what appear to be flypaper effects are in fact due to biases in evaluation methods.

Armed with data that allows a rigorous examination of this issue in the context of project aid, we first ascertained that full fungibility is possible, in that the bulk of the project communes would have spent more on road rehabilitation in the absence of the project than they in fact received from the project. We then used a difference-in-difference estimator coupled with propensity-score matching and weighting methods to determine the impacts of the aid. We find evidence of a flypaper effect of project aid on roads: spending on rehabilitation plus building of new roads accords reasonably closely to the total project allocation.

So we find evidence of behavioral responses by implementing agents away from donor-imposed restrictions. However, project aid resources stuck to the road sector as evidenced by the estimated kilometers of rehabilitated and built roads and by supporting evidence that the project did not lead to the construction of other basic infrastructure.

While impact evaluations have become more sophisticated in recent years — worrying about problems of selection and attrition bias, the heterogeneity of impacts and

so on — it remains rare for evaluations to establish that an intervention actually funded what it intended and was supplemental to local spending. It is our conclusion from this analysis that this should be ascertained as a first step in evaluating the impacts of development projects.

## References

- Alberto, Abadie and Guido Imbens (2006a), “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, 74(1), 235-267.
- Alberto, Abadie and Guido Imbens (2006b), “On the Failure of the Bootstrap for Matching Estimators,” mimeo, University of California, Berkeley.
- Feyzioglu, Tarhan, Vinaya Swaroop and Min Zhu (1998), “A Panel Data Analysis of the Fungibility of Foreign Aid,” *World Bank Economic Review*, 12(1), 29-58.
- Heckman, James and Petra Todd (1995), “Adapting Propensity Score Matching and Selection Models to Choice-Based Samples,” Working Paper, Department of Economics, University of Chicago.
- Hines, James and Richard Thaler (1995), “The Flypaper Effect,” *Journal of Economic Perspectives*, 9, 217-26
- Hirano, Keisuke and Guido Imbens, 2002, “Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization,” *Health Services and Outcomes Research Methodology*, 2, 259-278.
- Hirano, Keisuke, Guido Imbens and Geert Ridder (2003), “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161-1189.
- Imbens, Guido, 2004, “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *Review of Economics and Statistics*, 86(1), 4-29.
- Jacoby, Hanan (2002), “Is There an Intrahousehold ‘Flypaper Effect’? Evidence from a School Feeding Programme,” *The Economic Journal*, 112, 196-221.
- Khilji, Nasir and Ernest Zampelli (1994), “The Fungibility of U.S. Military and Non-military Assistance and the Impacts on Expenditures of Major Aid Recipients,” *Journal of Development Economics*, 43, 345-362.
- Pack, Howard and Janet Rothenberg Pack (1990), “Is Foreign Aid Fungible? The Case of Indonesia,” *The Economic Journal*, 100, 188-194.

- Pack, Howard and Janet Rothenberg Pack (1993), "Foreign Aid and the Question of Fungibility," *The Review of Economics and Statistics*, 75(2), 258-265.
- Rao, Govinda, Richard Bird and Jennie Litvack (1999), "The Changing Requirements of Fiscal Relations: Fiscal Decentralization in a Unified State," in Jennie Litvack and Dennis Rondinelli (eds.) *Market Reform in Vietnam: Building Institutions for Development*, Westport, Connecticut: Quorum Books.
- Ravallion, Martin (2006), "Evaluating Anti-Poverty Programs," in Robert Evenson and T. Paul Schultz (eds.) *Handbook of Development Economics*, Vol. 4, North-Holland, forthcoming.
- Roemer, John and Joaquim Silvestre (2002), "The 'Flypaper Effect' is Not an Anomaly," *Journal of Public Economic Theory*, 4(1), 1-17.
- Rosenbaum, Paul and Donald Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.
- Smith, A. Jeffrey and Petra E. Todd (2005), "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125, 305-353.
- van de Walle, Dominique (2007). "Evaluating the impacts of rural road projects," mimeo, DECRG, World Bank, Washington D.C.
- World Bank (1996), "Rural Transport Project Staff Appraisal Report," No. 15537-VN, East Asia and Pacific Region, World Bank, Washington, D.C.
- World Bank (1998), *Assessing Aid: What Works, What Doesn't, and Why?* Oxford University Press: New York.

**Table 1: Mean rural road kilometers, built and rehabilitated by communes pre- and post-project**

	Three years prior to project		Project period (four years)	
	Project	Non-project	Project	Non-project
<b>Total mean km of rural roads in commune</b>	32.11*	26.33	37.30*	31.52
<b>Total mean km built &amp; rehabilitated</b>	7.33	6.99	22.05**	17.17
<b>Mean km of new rural roads built</b>	2.70	2.29	7.26**	4.83
<i>with following road composition (%):</i>				
Paved, all weather roads	5.98	9.04	25.20	26.47
Paved, sometimes impassable	1.96	1.66	3.61	2.10
Earth road, motor vehicle passable	70.94	65.76	52.07	42.79
Earth road, motor vehicle impassable	21.12	23.54	19.12	28.65
<b>Mean km of rural roads rehabilitated</b>	4.62	4.70	14.79	12.34
<i>with following road composition (%):</i>				
Paved, all weather roads	7.29	9.96	27.42	20.31
Paved, sometimes impassable	2.76	2.40	3.82	1.23
Earth road, motor vehicle passable	59.75	64.10	54.65	60.38
Earth road, motor vehicle impassable	30.20	23.54	14.12	18.08

Note: Total sample has 194 observations, with 98 project and 96 comparison communes. \*\* and \* indicate single difference between project and non-project communes is significant at the 5 and 10% levels.



**Table 2: Logit regression of commune participation in RTP1**

	Coeff.	s.e.
Terrain: Coast	<i>Reference</i>	
Mountains	-0.56	1.06
Upland	0.04	0.82
Plains	-1.28	1.17
Province: Tra Vinh	<i>Reference</i>	
Lao Cai	1.80	1.08*
Thai Nguyen	2.15	1.30*
Nghe An	3.19	1.32**
Binh Thuan	2.69	1.41*
Kon Tum	4.04	1.22**
Population (log)	1.16	0.53**
Population density (log)	0.33	0.34
Minority population share	1.60	0.96*
National road passes through commune	-2.06	0.44**
Railway passes through commune without stop	1.63	0.65**
Waterway passes through commune	0.66	0.75
Distance to province center (km) (log)	-0.51	0.28*
Commune has a passenger transport service	-0.92	0.41**
Share of households engaged in non-agricultural activities	-3.02	2.16
Share of population working in government	-0.34	0.30
Share of population working in private enterprises	-0.29	0.13**
Share of population working in state enterprises	0.29	0.69
Share of crop land	1.71	1.67
Share of perennial crop land	2.91	3.50
Land rental market exists in commune	0.77	0.49
Number of production organizations	0.02	0.07
Commune has a radio broadcasting station	-0.53	0.47
Commune has a market	0.18	0.42
Agricultural crop land adversely affected by natural disaster (1996)	0.49	0.52
Commune has an agricultural bank	-1.15	0.61*
Number of official credit sources	-0.32	0.21
Enrollment rate for children age 6 to 15	-3.10	1.25**
Commune has a lower secondary school	0.03	0.60
Predicted consumption per capita (log)	1.35	1.09
Share of households owning motorcycles	0.06	0.04
Road density (commune and district level roads)	-16.68	9.46*
Share of earth and car impassable roads in total road km	1.17	0.75
Constant	-18.69	8.27**
Number of Observations	200	
Pseudo R <sup>2</sup>	0.25	

Note: (i) The dependent variable equals one for project communes and zero for non-project communes. (ii) Standard errors are clustered at the district level of which there are 29. (iii) \*\* and \* indicate significance levels of 5 and 10%. (iv) Average commune household consumption is predicted using a consumption model calibrated to the 1998 VLSS.

**Table 3: Double difference estimates of rural road kilometer increments during project period (total km)**

	Total sample			Treatment support			Common support		
	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD
<b>Total km</b>	4.88 (1.68)**	5.76 (2.45)**	6.76 (2.39)**	4.64 (1.70)**	5.75 (2.30)**	6.76 (2.39)**	3.77 (1.51)**	4.57 (2.27)**	5.88 (2.29)**
<i>Of which:</i>									
<b>Built</b>	2.43 (1.04)**	2.77 (1.37)**	2.89 (0.94)**	2.42 (1.04)**	2.76 (1.17)**	2.89 (0.94)**	2.09 (1.02)**	2.38 (1.03)**	2.56 (0.92)**
<b>Rehabilitated</b>	2.45 (1.48)	3.00 (1.87)	3.88 (2.13)*	2.23 (1.46)	2.99 (1.85)	3.87 (2.14)*	1.68 (1.38)	2.19 (1.89)	3.32 (2.10)
<b>Allocated by RTP1</b> (Project database)		4.58			4.58			4.50	
<b>p-value for test:</b>									
rehab=allocated	0.16	0.40	0.74	0.12	0.40	0.74	0.05	0.23	0.32
total=allocated	0.86	0.64	0.37	0.97	0.60	0.37	0.60	0.98	0.58

Note: (i) Standard errors clustered at the district level are given in parentheses; PS kernel matched standard errors are obtained by bootstrapping (100 repetitions). (ii) \*\* and \* indicate significance levels of 5 and 10%; (iii) The total sample has 194 observations (96 comparison and 98 project communes); treatment support has 191 observations (93 comparison and 98 project communes); common support has 185 observations (93 comparison and 92 project communes).

**Table 4a: Communes that rehabilitated less pre-project than RTP1 allocation: Double difference estimates of project impact on rural road kilometer increments (total km)**

	Total sample			Treatment support			Common support		
	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD
<b>Total km</b>	6.26 (2.63)**	7.20 (2.98)**	8.14 (3.36)**	5.32 (2.73)*	7.08 (2.94)**	8.13 (3.41)**	4.35 (2.26)*	5.91 (2.73)**	7.15 (3.08)**
<i>Of which:</i>									
<b>Built</b>	2.41 (1.37)**	2.78 (1.56)*	2.87 (1.32)**	3.00 (1.38)**	2.80 (1.62)*	2.94 (1.33)**	2.50 (1.24)**	2.43 (1.40)	2.26 (1.13)**
<b>Rehabilitated</b>	3.85 (1.91)**	4.42 (2.22)**	5.28 (2.54)**	2.32 (1.89)	4.28 (2.18)**	5.19 (2.59)**	1.84 (1.71)	3.64 (2.09)*	4.72 (2.48)*
<b>Allocated by RTP1</b> (Project database)		5.67			5.67			5.55	
<b>p-value for test:</b>									
rehab=allocated	0.35	0.57	0.87	0.09	0.52	0.85	0.03	0.34	0.70
total=allocated	0.82	0.61	0.46	0.89	0.63	0.48	0.56	0.93	0.63

Note: (i) Standard errors clustered at the district level are given in parentheses; PS kernel matched standard errors are obtained by bootstrapping (100 repetitions). (ii) \*\* and \* indicate significance levels of 5 and 10%; (iii) The total sample has 149 observations (96 comparison and 53 project communes); treatment support has 119 observations (66 comparison and 53 project communes); common support has 114 observations ( 66 comparison and 51 project communes). (iv) The sample of project communes for this table are those that rehabilitated less km pre-project than were allocated by RTP1.

**Table 4b: Communes that rehabilitated more pre-project than RTP1 allocation: Double difference estimates of project impact on rural road kilometer increments (total km)**

	Total sample			Treatment support			Common support		
	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD
<b>Total km</b>	3.26 (1.78)*	4.07 (2.23)*	5.14 (2.09)**	3.02 (1.78)*	4.04 (2.60)**	5.13 (2.09)**	2.17 (1.63)	2.73 (2.43)	4.28 (1.96)**
<i>Of which:</i>									
<b>Built</b>	2.45 (1.41)*	2.75 (1.24)**	2.91 (1.31)**	2.44 (1.41)*	2.74 (1.29)**	2.91 (1.31)**	2.33 (1.47)	2.54 (1.28)**	2.80 (1.35)**
<b>Rehabilitated</b>	0.80 (1.60)	1.32 (1.77)	2.23 (2.10)	0.58 (1.54)	1.30 (2.13)	2.22 (2.10)	-0.16 (1.49)	0.20 (1.93)	1.48 (2.07)
<b>Allocated by RTP1</b> (Project database)		3.30			3.19			3.19	
<b>p-value for test:</b>									
rehab=allocated	0.13	0.27	0.61	0.09	0.36	0.61	0.03	0.13	0.42
total=allocated	0.98	0.73	0.38	0.87	0.38	0.77	0.48	0.58	0.82

Note: (i) Standard errors clustered at the district level are given in parentheses; PS kernel matched standard errors are obtained by bootstrapping (100 repetitions). (ii) \*\* and \* indicate significance levels of 5 and 10%; (iii) The total sample has 141 observations (96 comparison and 45 project communes); treatment support has 138 observations (93 comparison and 45 project communes); common support has 134 observations (93 comparison and 41 project communes). (iv) The sample of project communes for this table are those that rehabilitated more km pre-project than were allocated by RTP1.

**Table 5: Double difference estimates of whether different types of non-road infrastructure building occurred during the project period**

	Total sample			Treatment support			Common support		
	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD
Schools	0.08 (0.06)	0.16 (0.10)	0.24 (0.16)	0.08 (0.06)	0.16 (0.11)	0.24 (0.16)	0.09 (0.06)	0.14 (0.09)	0.24 (0.16)
Health center	-0.02 (0.08)	0.12 (0.09)	0.18 (0.09)*	-0.02 (0.09)	0.12 (0.11)	0.18 (0.09)*	-0.05 (0.09)	0.07 (0.11)	0.14 (0.10)
Commune cultural center	-0.06 (0.05)	-0.17 (0.09)*	-0.14 (0.10)	-0.07 (0.05)	-0.17 (0.11)*	-0.14 (0.10)	-0.06 (0.05)	-0.16 (0.10)	-0.14 (0.10)
Commune offices	0.03 (0.08)	0.09 (0.10)	0.14 (0.11)	0.03 (0.08)	0.09 (0.10)	0.14 (0.11)	0.02 (0.09)	0.05 (0.11)	0.12 (0.11)
Any construction	0.04 (0.04)	0.15 (0.09)*	0.24 (0.16)	0.05 (0.04)	0.15 (0.10)	0.24 (0.16)	0.04 (0.04)	0.12 (0.09)	0.24 (0.16)

Notes: (i) The impact variable is a dummy variable for whether the commune did construction on the various types of infrastructure. (ii) Standard errors clustered at the district level are given in parentheses; (iii) \*\* and \* indicate significance levels of 5 and 10%. (iv) The total sample has 194 observations (96 comparison and 98 project communes); treatment support has 191 observations (93 comparison and 98 project communes); common support has 185 observations (93 comparison and 92 project communes).

**Table 6: Double difference estimates of increments to different types of rural roads during project period (total km)**

	Total sample			Treatment support			Common support		
	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD	DD	PS kernel matched DD	PS weighted DD
<b>Road building</b>									
Paved, all-weather	-0.06 (0.35)	0.25 (0.28)	0.32 (0.32)	-0.09 (0.36)	0.25 (0.28)	0.32 (0.33)	-0.06 (0.36)	0.25 (0.28)	0.35 (0.33)
Paved, sometimes impassable	0.14 (0.09)	0.16 (0.10)	0.16 (0.09)	0.14 (0.08)	0.16 (0.10)	0.16 (0.09)	0.15 (0.09)	0.17 (0.10)*	0.17 (0.10)
Earth, all-weather	1.50 (0.80)*	1.73 (1.03)*	2.03 (0.67)**	1.48 (0.82)*	1.73 (0.91)*	2.03 (0.67)**	1.12 (0.76)	1.28 (1.02)	1.66 (0.58)**
Earth, sometimes impassable	-0.08 (0.34)	0.44 (0.36)	0.49 (0.41)	-0.001 (0.33)	0.44 (0.38)	0.49 (0.41)	0.04 (0.34)	0.45 (0.40)	0.53 (0.42)
<b>Road rehabilitation</b>									
Paved, all-weather	1.19 (0.56)**	1.30 (0.64)**	1.49 (0.57)**	1.23 (0.55)**	1.31 (0.67)**	1.49 (0.57)**	1.12 (0.58)*	1.11 (0.59)*	1.38 (0.59)**
Paved, sometimes impassable	0.31 (0.21)	0.40 (0.16)**	0.40 (0.19)**	0.30 (0.21)	0.40 (0.16)**	0.40 (0.19)**	.33 (0.22)	0.42 (0.18)**	0.43 (0.20)**
Earth, all-weather	1.16 (1.21)	2.35 (1.26)*	2.99 (1.35)**	1.00 (1.22)	2.35 (1.37)*	2.99 (1.35)**	0.87 (1.24)	2.00 (1.39)	2.85 (1.35)**
Earth, sometimes impassable	0.31 (1.06)	0.33 (1.09)	0.61 (1.21)	0.22 (1.08)	0.33 (1.07)	0.61 (1.21)	0.01 (1.06)	0.11 (1.31)	0.39 (1.20)

Note: (i) Standard errors clustered at the district level are given in parentheses; (ii) \*\* and \* indicate significance levels of 5 and 10%. (iii) The total sample has 194 observations (96 comparison and 98 project communes); treatment support has 191 observations (93 comparison and 98 project communes); common support has 185 observations (93 comparison and 92 project communes).

Figure 1. Distribution of estimated propensity scores

